(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁷:	G06F 17/30	(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(21) International Application Number: PCT/US00/42832

(22) International Filing Date:
22 December 2000 (22.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/469,567   22 December 1999 (22.12.1999)   US

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(71) Applicants and
(72) Inventors: PACE, Mark, Raymond [US/US]; 42 15th Avenue, San Mateo, CA 94402 (US). TALLEY, Brooks, Cash [US/US]; 42 15th Avenue, San Mateo, CA 94402 (US).

(74) Agents: VIERRA, Larry, E. et al.; Fliesler Dubb Meyer and Lovejoy LLP, Four Embarcadero Center, Suite 400, San Francisco, CA 94111-4156 (US).

Published:
— With international search report.
— Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DISTRIBUTED CONTENT IDENTIFICATION SYSTEM

(57) Abstract: An e-mail sender (10) transfers an e-mail to its intended recipient (40). The message arrives at a first tier system (20) which may be an e-mail server. A digital identifier engine on the first tier system (20) generates a digital identifier which comprises a hash of at least a portion of the e-mail. Second tier system (30) includes a database and processor which determines, based on an algorithm which varies with the characteristics tested, whether the e-mail is spam or not processes the e-mail accordingly.

DISTRIBUTED CONTENT IDENTIFICATION SYSTEM

BACKGROUND OF THE INVENTION

Field of the Invention

The invention relates to the field of content identification for files on a network.

5

Description of the Related Art

With the proliferation and growth of the Internet, content transfer between systems on both public and private networks has increased exponentially. While the Internet has brought a good deal of information to a large number of people in a

10   relatively inexpensive manner, this proliferation has certain downsides. One such downside, associated with the growth of e-mail in particular, is generally referred to as "spam" e-mail. Spam e-mail is unsolicited e-mail which is usually sent out in large volumes over a short period of time with the intent of inducing the recipient into availing themselves of sales opportunities or "get rich quick" schemes.

15   To rid themselves of spam, users may resort to a number of techniques. The most common is simple filtering using e-mail filtering which is built into e-mail client programs. In this type of filtering, the user will set up filters based on specific words, subject lines, source addresses, senders or other variables, and the e-mail client will process the incoming e-mail when it is received, or at the server level, and

20   take some action depending upon the manner in which the filter is defined.

More elaborate e-mail filtering services have been established where, for a nominal fee, off-site filtering will be performed at a remote site. In one system, e-mails are forwarded offsite to a service provider and the automatic filtering occurs at the provider's location based on heuristics which are updated by the service

25   provider. In other systems, offsite filtering occurs using actual people to read through e-mails and judge whether e-mail is spam or not. Other systems are hybrids, where heuristics are used and, periodically, real people review e-mails which are forwarded to the service to determine whether the e-mail constitutes "spam" within the aforementioned definition. In these hybrid services, personal reviews occur on

30   a random basis and hence constitute only a spot check of the entire volume of e-mail which is received by the service. In systems where real people review e-mails, confidentiality issues arise since e-mails are reviewed by a third party who may or may not be under an obligation of confidentiality to the sender or recipient of the e-

mail.

In addition, forwarding the entire e-mail including attachments to an outside service represents a high bandwidth issue since effectively this increases the bandwidth for a particular e-mail by three times: once for the initial transmission, the second time for the transmission to the service and the third time from the service back to server for redistribution to the ultimate recipient.

Further, senders of spam have become much more sophisticated at avoiding the aforementioned filters. The use of dynamic addressing schemes, very long-length subject lines and anonymous re-routing services makes it increasingly difficult for normal filtering schemes, and even the heuristics-based services discussed above, to remain constantly up-to-date with respect to the spammers' ever changing methods.

Another downside to the proliferation of the Internet is that it is a very efficient mechanism for delivering computer viruses to a great number of people. Virus identification is generally limited to programs which run and reside on the individual computer or server in a particular enterprise and which regularly scan files and e-mail attachments for known viruses using a number of techniques.

SUMMARY OF THE INVENTION

Hence, the object of the invention is to provide a content classification system which identifies content in an efficient, up-to-date manner.

The further object of the invention is to leverage the content received by other users of the classification system to determine the characteristic of the content.

Another object of the invention is to provide a service which quickly and efficiently identifies a characteristic of the content of a given transmission on a network at the request of the recipient.

Another object of the invention is to provide the above objects in a confidential manner.

A still further object is to provide a system which operates with low bandwidth.

These and other objects of the invention are provided in the present invention. The invention, roughly described, comprises a file content classification system. In one aspect the system includes a digital ID generator and an ID database coupled to receive IDs from the ID generator. The system further includes a characteristic comparison routine identifying the file as having a characteristic based

on ID appearance in the appearance database.

In a particular embodiment, the file is an e-mail file and the system utilizes a hashing process to produce digital IDs. The IDs are forwarded to a processor via a network. The processor performs the characterization and determination steps. The

5   processor then replies to the generator to enable further processing of the email based on the characterization reply.

In a further aspect, the invention comprises a method for identifying a characteristic of a data file. The method comprises the steps of: generating a digital identifier for the data file and forwarding the identifier to a processing system;

10   determining whether the forwarded identifier matches a characteristic of other identifiers; and processing the e-mail based on said step of determination.

In yet another aspect, the invention comprises a method for providing a service on the Internet, comprising: collecting data from a plurality of systems having a client agent on the Internet to a server having a database; characterizing the data received relative to

15   information collected in the database; and transmitting a content identifier to the client agent. In this aspect, said step of collecting comprises collecting a digital identifier for a data file. In addition, said step of characterizing comprises: tracking the frequency of the collection of a particular identifier; characterizing the data file based on said frequency; storing the characterization; and comparing collected identifiers to the known

20   characterization

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be described with respect to the particular embodiments thereof. Other objects, features, and advantages of the invention will become apparent with

25   reference to the specification and drawings in which:

Figure 1 is a block diagram indicating the system in filtering e-mail to identify content in accordance with the prior art.

Figure 2 is a block diagram illustrating the process of the present invention.

Figure 3 is a block diagram illustrating in additional detail the method and

30   apparatus of the present invention.

Figure 4 is a block diagram illustrating a second embodiment of the method and apparatus of the present invention.

DETAILED DESCRIPTION

35   The present invention provides a distributed content classification system which

utilizes a digital identifier for each piece of content which is sought to be classified, and characterizes the content based on this ID. In one aspect of the system, the digital identifier is forwarded to a processing system which correlates any number of other identifiers through a processing algorithm to determine whether a particular characteristic

5   for the content exists. In essence, the classification is a true/false test for the content based on the query for which the classification is sought. For example, a system can identify whether a piece of e-mail is or is not spam, or whether the content in a particular file matches a given criteria indicating it is or is not copyrighted material or contains or does not contain a virus.

10   While the present invention will be discussed with respect to classifying e-mail messages, it will be understood by those of average skill in the art that the data classification system of the present invention can be utilized to classify any sort of text or binary data which resides on or is transmitted through a system.

       Figure 1 is a high level depiction of the present invention wherein an e-mail sender
15   10 transmits an e-mail which is intercepted by a filtering process/system 15 before being forwarded to the sender. The system has the ability to act on the e-mail before the recipient 20 ever sees the message.

       Figure 2 illustrates the general process of the present invention in the e-mail context when an e-mail sender 10 transfers an e-mail to its intended recipient 40, the
20   message arrives at a first tier system 20 which in this example may represent an e-mail server. Normally (in the absence of the system of the present invention), the first tier system 20 will transmit an e-mail directly to the intended recipient when the recipient's e-mail client application requests transmission of the e-mail. In the present invention, a digital identifier engine on the first tier system cooperating with the e-mail server will generate a
25   digital identifier which comprises, in one environment, a hash of at least a portion of the e-mail. The digital identifier is then forwarded to a second tier system 30. Second tier system 30 includes a database and processor which determines, based on an algorithm which varies with the characteristic tested, whether the e-mail meets the classification of the query (e.g. is it spam or not?).

30   Based on the outcome of this algorithm, a reply is sent from the second tier system 30 to the first tier system 20, where the system then processes the e-mail in accordance with the regenerated description by the user based on the outcome of the filter. The result can be as shown in Figure 2, the filtered e-mail product being forwarded to the e-mail recipient. Other options for disposition of the e-mail depending upon the outcome of the
35   algorithm computed at second tier system 30 are described below.

It should be understood with reference to Figures 1 and 2 that the external e-mail sender can be any source of electronic mail or electronic data sent to the filtering process from sources outside the system. The e-mail recipients 40 represent the final destination of electronic data that passes through the filtering process.

5   In one aspect, the system may be implemented in executable code which runs on first tier system 20 and generates digital IDs in accordance with the MD5 hash fully described at http://www.w3.org/TR/1998/Rec-DSig-label/MD5_1_6. It should be recognized however that any hashing algorithm can be utilized. In one embodiment, the digital ID generated by the MD5 hash is of the entire subject line up to the point where two

10   spaces appear, the entire body, and the last 500 bytes of the body of the message. It should be further understood that the digital ID generated may be one hash, or multiple hashes, and the hashing algorithm may be performed on all or some portion of the data under consideration. For example, the hash may be of the subject line, some number of characters of the subject line, all of the body or portions of the body of the message. It

15   should further be recognized that the digital ID is not required to be of fixed length.

The first tier executable may be run as a separate process or as a plug-in with the e-mail system running on a first tier system 20. In one embodiment, the executable interfaces with a commonly used mail server on a running system such as a first tier system 20 is known as Sendmail™. A common set of tools utilized with Sendmail™ is

20   Procmail. (http://www.ii.com/internet/robots/procmail). In one aspect of the system of the present invention, the executable may interface with Sendmail™ and Procmail. In such an embodiment, a configuration file (such as a sendmail.cf) includes a line of code which instructs the Procmail server program to process incoming e-mails through the first tier site e-mail executable to generate and transport digital IDs to the second tier system, receive

25   its reply, and instruct the Procmail to process or delete the message, as a result of the reply message.

It should be understood that the executable may be written in, for example, perl script and can be designed to interact with any number of commercial or free e-mail systems, or other data transfer systems in applications other than e-mail.

30   The digital ID usage in this context reduces bandwidth which is required to be transported across the network to the second tier system. Typically, the ID will not only contain the hashed data, but may include versioning information which informs the second tier system 30 of the type of executable running on the first tier system 20.

In addition, the reply of the second tier system to the first tier system may be, for

35   example, a refusal of service from the second tier system 30 to the first tier system 20 in

cases where the first tier system is not authorized to make such requests. It will be recognized that revenue may be generated in accordance with the present invention by providing the filtering service (i.e. running the second tier service process and maintaining the second tier database) for a fee based on volume or other revenue criteria. In this

5   commercial context, the reply may be a refusal of service of the user of the first tier system 20 which has exceeded their allotted filtering quota for a given period.

Figure 3 shows a second embodiment of the system of the present invention. In Figure 3, the first tier system is broken down into three components including a message preprocessing section 110, a message processing section 120, a configuration file DS10.

10   In this example, the e-mail from sender 10 is first diverted to message preprocessing 110. Preprocessing algorithm is configured with rules from configuration file DS10. These rules are guidelines on how and when, for example, to generate digital IDs from the e-mail which is received. Message preprocessing receives the email from the e-mail sender 10 and generates digital IDs based on the preprocessing rules from DS10. DS10 is a configuration

15   file which stores configuration rules (before preprocessing and postprocessing) for the first tier system 20. The message processing rules may include guidelines on how to dispose of those e-mails classified as spam. For example, a message may be detected, and may be forwarded to a holding area for electronic mail that has been deemed to be spam by second tier system 30, have the word "SPAM" added to the subject line, moved to a

20   separate folder, and the like. In this example, message preprocessing rules include rules which might exempt all e-mails from a particular destination or address from filtering by the system. If a message meets such exemption criteria, the message is automatically forwarded, as shown on line 50, directly to message processing 120 for forwarding directly onto the e-mail recipient 40. Such rules may also comprise criteria for forwarding an e-mail

25   directly to a rejected message depository DS20.

If a preprocessing rule does not indicate a direct passage of a particular e-mail through the system, one or more digital identifiers will be generated as shown at line 66 and transmitted to the second tier system 30'. In the example shown at Figure 3, second tier system 30' includes a second tier server 210 in a third tier database 220. In this

30   example, the second tier server relays digital IDs and replies between preprocessing and the message processing 120. The example shown in Figure 3 is particularly useful in an Internet based environment where the second tier server 210 may comprise a web server which is accessible through the Internet and the third tier database 220 is shielded from the Internet by the second tier server through a series of firewalls or other security measures.

35   This ensures that the database of digital ID information which is compiled at the third tier

database 220 is free from attack from individuals desirous of compromising the security of this system.

In this case, second tier server 210 forwards the digital ID directly to the third tier database which processes the IDs based on the algorithm for testing the data in question.
5   The third tier database generates a reply which is forwarded by the second tier server back to message processing 120. Message processor 120 can then act on the e-mail by either sending filtered e-mail to the e-mail recipient, sending the filtered e-mail to the rejected message depository DS20 or acting on the message in accordance with user-chosen configuration settings specified in configuration file DS10.

10   In the environment shown in Figure 3, the configuration file DS110 on the first tier allows other decisions about the e-mail received from the e-mail sender 10 to be made, based on the reply from second tier 30. For example, in addition to deleting spam e-mail, the subject line may be appended to indicate that the e-mail is "spam," the e-mail may be held in a quarantine zone for some period of time, an auto reply generated, and the like.
15   In addition, the message preprocessing and message processing rules allow decisions on e-mail processing to account for situations where second tier system 30' is inaccessible. Decisions which may be implemented in such cases may include "forward all e-mails," "forward no e-mails," "hold for further processing," and the like.

In an Internet based environment, the second tier server 30 may transmit a digital
20   identification and other information to the third tier database 220 by means of the HTTP protocol. It should be recognized that other protocols may be used in accordance with the present invention. The third tier database 220 may be maintained on any number of different commercial database platforms. In addition the third tier database may include system management information, such as client identifier tracking, and revenue processing
25   information. In an unique aspect of the present invention in general, the digital IDs in third tier database 22 are maintained on a global basis. That is, all first tier servers which send digital IDs to second tier servers 210 contribute data to the database and the processing algorithm running on the third tier system. In one embodiment, where spam determination is the goal, the algorithm computes, for example, the frequency with which a message (or,
30   in actuality, the ID for the message), is received within a particular time frame. For example, if a particular ID indicating the same message is seen some number of times per hour, the system classifies the message (and ID) as spam. All subsequent IDs matching the ID classified as spam will now cause the system 30' to generate a reply that the e-mail is spam. Each client having a first tier system 20' which participates in the system of the
35   present invention benefits from the data generated by other clients. Thus, for example, if

a particular client receives a number of spam e-mails meeting the frequency requirement causing the system to classify another client having a first tier system 20' which then sees a similar message will automatically receive a reply that the message is spam.

5   It should be recognized that in certain cases, large reputable companies forward a large block of e-mails to a widespread number of users, such as, for example information mailing list servers specifically requested by e-mail receivers. The system accounts for such mailing list application on both the area and second tier system levels. Exceptions may be made in the algorithm running on the third tier database 220 to take into account the fact that reputable servers should be allowed to send a large number of e-mails to a

10   large number of recipients at the destination system 20'. Alternatively, or in conjunction with such exceptions, users may define their own exceptions via the DS10 configuration. As a service, any number of acceptable sources such as, for example, the Fortune 1000 companies' domain names may be characterized as exempted "no spam" sites, and users can choose to "trust" or "not trust" server side settings.

15   While the aforementioned embodiment utilizes a frequency algorithm to determine whether a message is spam, additional embodiments in the algorithm can analyze messages for the frequency of particular letters or words, and/or the relationship of the most common words to the second most common words in a particular message. Any number of variants of the algorithm may be used.

20   It should be further recognized that the second tier server can be utilized to interface with the value added services, such as connecting the users to additional mailing lists and reference sources, providing feedback on the recipients' characteristics to others, and the like.

Figure 4 shows a further embodiment of the invention and details how the server

25   side system manipulates with the digital identifiers. In Figure 4, the embodiment includes a DS10 configuration file which provides message exemption criteria and message processing rules to both message preprocessing and routine 110 message processing 120, respectively. Message preprocessing 110 may be considered as two components: message exemption checking 111 and digital ID creation 112. Both of these components

30   function as described above with respect to Figure 3 allowing for exempt e-mails to be passed directly to an e-mail recipient 40, or determining whether digital IDs need to be forwarded to second tier server 210. Replies are received by message processing algorithm 120 is acted on by rule determination algorithm 121, and e-mail filtering 123.

At the second tier system 30', digital IDs transmitted from second tier processor

35   210 are transmitted to a digital ID processor 221. In this embodiment, processor 221

increments counter data stored in DS30 for each digital ID per unit time. As the volume of messages processed by database 220 can be quite large, the frequency algorithm may be adjusted to recognize changes in the volume of individual messages seen as a percentage of the total message volume of the system.

5   The frequency data stored at DS30 feeds a reply generator 222 which determines, based on both the data in the DS30 and particular information for a given client, (shown as data record DS40) whether the reply generated and forwarded to second tier server 210 should indicate that the message is spam or not. Configuration file DS40 may include rules, as set forth above, indicating that the reply from the second tier server 210 is 10   forwarded to rule determination component of message processor 120 which decides, as set forth above, how to process the rule if it is in fact determined that it is spam. The filtered e-mail distribution algorithm forwards the e-mail directly to the e-mail client 40 or to the rejected message repository as set forth above.

        A key feature of the present invention is that the digital IDs utilized in the data 15   identifier repository DS30 are drawn from a number of different first tier systems. Thus, the greater number of first tier systems which are coupled to the second tier server and subsequent database 220, the more powerful the system becomes.

        It should be further recognized that other applications besides the detection of spam e-mail include the detection of viruses, and the identification of copyrighted material 20   which are transmitted via the network. Moreover, it should be recognized that the algorithm for processing digital identifiers and the data store DS30 are not static, but can be adjusted to look for other characteristics of the message or data which is being tested besides frequency.

        Hence, the system allows for leveraging between the number of first tier systems 25   or clients coupled to the database to provide a filtering system which utilizes a limited amount of bandwidth while still providing a confidential and powerful e-mail filter. It should be further recognized that the maintainer of the second and third tier systems may generate revenue for the service provided by charging a fee for the service of providing the second tier system process.

30   Still further, the system can collect and distribute anonymous statistical data about the content classified. For example, where e-mail filtering is the main application of the system, the system can identify the percentage of total e-mail filtered which constitutes spam, where such e-mail originates, and the like, and distribute it to interested parties for a fee or other compensation.

35

CLAIMS

What is claimed is:

1.   A file content classification system comprising:

5   a digital ID generator;

an ID appearance database coupled to receive IDs from the ID generator; and

a characteristic comparison routine identifying the file as having a characteristic

based on ID appearance in the appearance database.

10   2.   The content classification system of claim 1 wherein said ID generator comprises

a hashing algorithm.

3.   The content classification system of claim 2 wherein said hashing algorithm is the

MD5 hashing algorithm.

15

4.   The content classification system of claim 1 wherein said ID appearance database

tracks the frequency of appearance of a digital ID.

5.   The content classification system of claim 1 further including a plurality of digital

20   ID generators on different systems all coupled to and providing IDs to said ID appearance

database.

6.   The content classification system of claim 5 wherein said plurality of digital ID

generators are coupled to said database via a combination of public and private networks.

25

7.   The content classification system of claim 6 wherein said database is coupled to

an intermediate server which is coupled to said plurality of generators.

8.   The content classification system of claim 6 wherein said intermediate server is a

30   web server.

9.   The content classification system of claim 1 wherein said characteristic comprises

junk e-mail and said characteristic is defined by a frequency of appearance of a digital ID.

35   10.   A method for identifying a characteristic of a data file, comprising:

generating a digital identifier for the data file and forwarding the identifier to a processing system;

determining whether the forwarded identifier matches a characteristic of other identifiers; and

5

processing the email based on said step of determining.

11.   The method of claim 10 wherein said step of generating comprises hashing at least a portion of the data file.

10   12.   The method of claim 11 wherein said step of hashing comprises using the MD5 hash.

13.   The method of claim 11 wherein said step of generating comprises hashing multiple portions of the data file.

15

14.   The method of claim 10 wherein said data file is an email message and said step of determining comprises determining whether said email is spam.

15.   The method of claim 10 wherein said step of determining identifies said e-mail as
20   spam by tracking the rate per unit time a digital ID is generated.

16.   The method of claim 10 wherein said step of generating comprises generating IDs at a plurality of source systems all coupled via a network to at least one processing system performing the determining step.

25

17.   The method of claim 16 wherein said step of processing comprises instructing said plurality of source systems to perform an action with the email based on said determining step.

30   18.   A method of filtering an email message, comprising:

processing the message to provide a digital identifier;

comparing the digital identifier to a characteristic database of digital identifiers to determine whether the message has said characteristic; and

processing the message based on said step of comparing.

35

19. The method of claim 18 wherein said step of processing occurs on at least one first system, and said step of comparing occurs on a second system.

20. The method of claim 19 wherein said step of processing occurs on a plurality of first systems.

21. The method of claim 19 wherein said at least one first system and second system are coupled by the Internet.

22. The method of claim 18 wherein said step of comparing comprises determining the frequency of a particular ID occurring in a time period, classifying said ID as having a characteristic, and comparing digital identifiers to said classified IDs.

23. A file content classification system, comprising:
a first system having a file to be classified;
an file ID generator on the fist system;
a database on a second system coupled to the ID generator to receive IDs generated by the ID generator;
a comparison routine on the second system classifying the ID relative to the database as meeting or not meeting a characteristic.

24. The system of claim 23 including a plurality of first systems each including a respective file ID generator coupled to the database on the second system.

25. The system of claim 24 wherein the plurality of first systems is coupled to the second system via the Internet.

26. The system of claim 25 wherein the second system comprises a web server interface system and a database system, wherein the database system is isolated from the Internet by the web server system.

27. A content classification system for a first and second computer coupled by a network, comprising:
a client agent file identifier generator on the first computer; and
a server comparison agent and data-structure on the second computer receiving

identifiers from the client agent and providing replies to the client agent;

      wherein the client agent processes the file based on replies from the server comparison agent.

5   28.   A method for providing a service on the Internet, comprising:

      collecting data from a plurality of systems having a client agent on the Internet to a server having a database;

      characterizing the data received relative to information collected in the database; and

10   transmitting a content identifier to the client agent.

      29.   The method of claim 28 wherein said step of collecting comprises collecting a digital identifier for a data file.

15   30.   The method of claim 28 wherein said data file is an e-mail.

      31.   The method of claim 29 wherein said step of characterizing comprises:

      tracking the frequency of the collection of a particular identifier;

      characterizing the data file based on said frequency;

20   storing the characterization; and

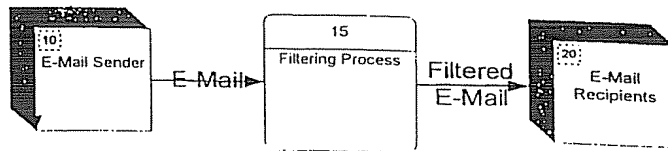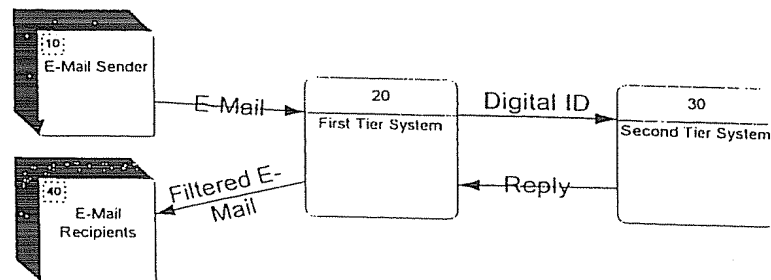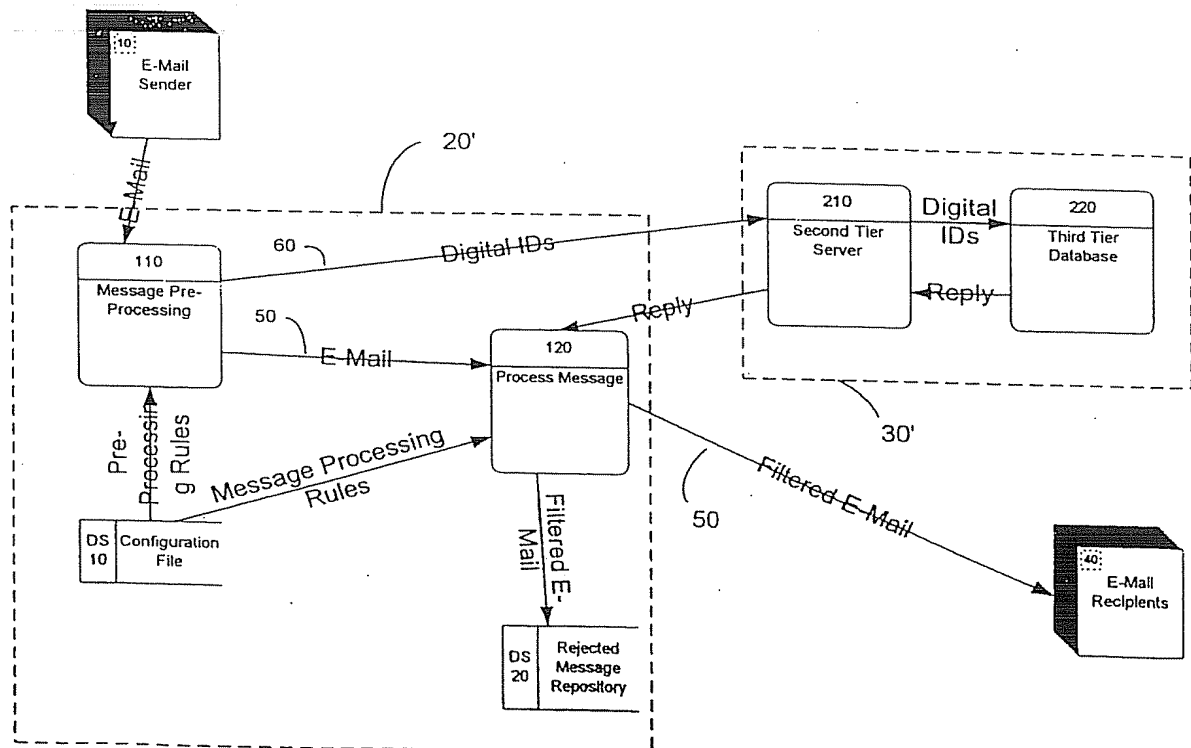   comparing collected identifiers to the known characterization.

Figure 1

10
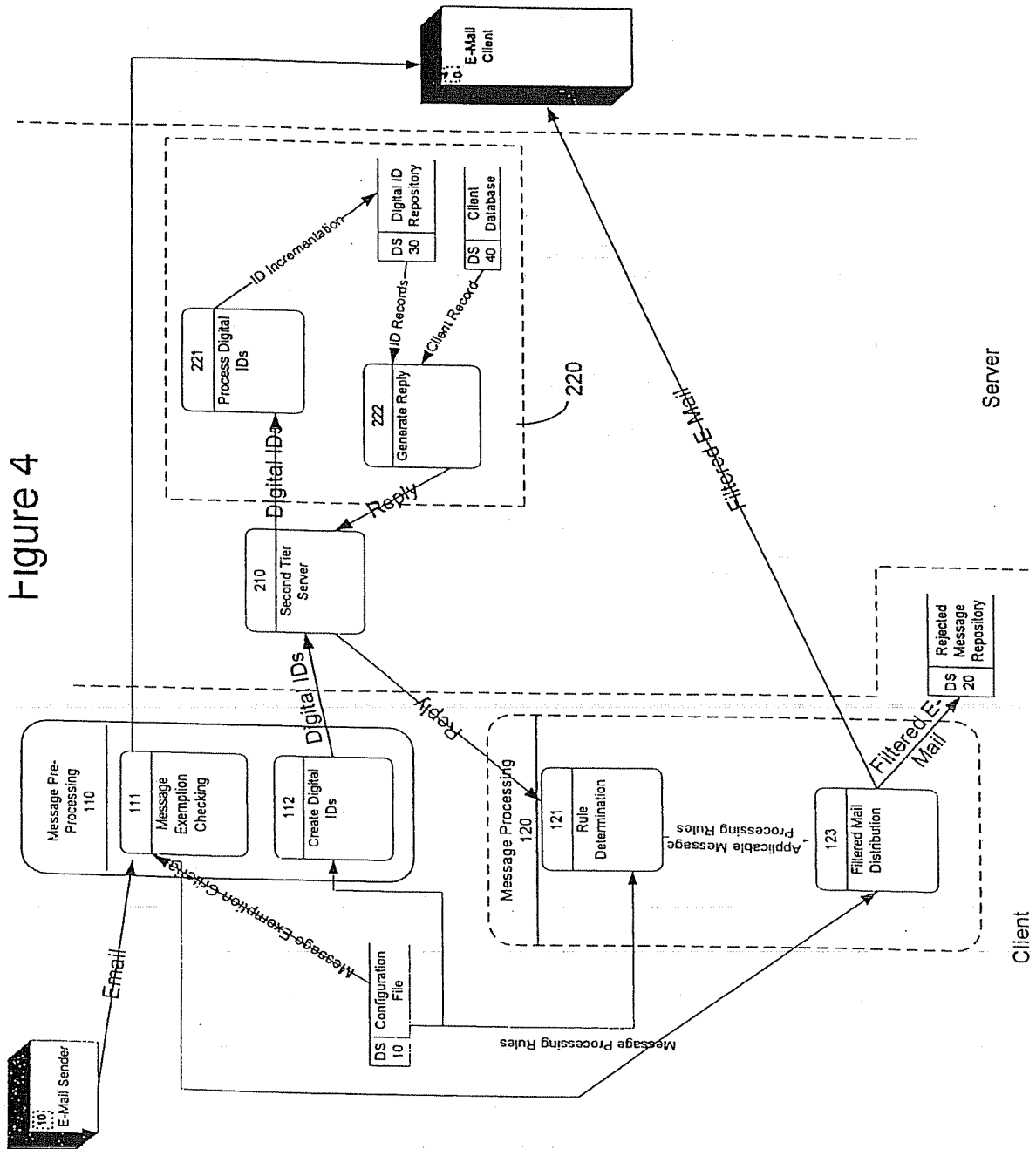E-Mail Sender

E-Mail→

15
Filtering Process

Filtered
E-Mail

20
E-Mail
Recipients

Figure 2

10
E-Mail Sender

E-Mail→

20
First Tier System

Digital ID→

30
Second Tier System

40
E-Mail
Recipients

←Filtered E-
Mail

←Reply

Figure 3

10
E-Mail
Sender

E-Mail

20'

110
Message Pre-
Processing

60
Digital IDs→

50
E-Mail→

120
Process Message

←Reply

210
Second Tier
Server

Digital
IDs→

←Reply

220
Third Tier
Database

30'

Pre-
Processing Rules

DS
10
Configuration
File

Message Processing
Rules

Filtered E-
Mail

DS
20
Rejected
Message
Repository

50

Filtered E-Mail

40
E-Mail
Recipients

Figure 4

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G06F 17/30

US CL : 707/9, 104; 709/202, 204

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 707/6, 9, 104; 709/201, 202, 204, 225

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EAST

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5,884,033 A (DUVALL et al) 16 March 1999 (16.03.1999), ALL.	1-31
X, P	US 6,052,709 A (PAUL) 18 April 2000 (18.04.2000), All.	1-31

☐ Further documents are listed in the continuation of Box C.   ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier application or patent published on or after the international filing date		
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 April 2001 (26.04.2001)	14 MAY 2001
Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks · Box PCT Washington, D.C. 20231	Uyen T Le *James R. Matthews*
Facsimile No. (703)305-3230	Telephone No. 305-9000

Form PCT/ISA/210 (second sheet) (July 1998)